# A RE-EXAMINATION OF C. E. M. HANSEL'S CRITICISM OF THE PRATT-WOODRUFF EXPERIMENT

By R. G. MEDHURST and CHRISTOPHER SCOTT[1]

**ABSTRACT: This paper attempts a reassessment of the evidence regarding the hypothesis advanced by C. E. M. Hansel to explain in normal terms the results of the Pratt-Woodruff experiment.**

**The statistical evidence in support of Hansel's hypothesis is examined for all significantly scoring subjects, using tests which are more appropriate than those employed either by Hansel or by Pratt and Woodruff. A significant effect of the type claimed by Hansel is found not only in the highest scoring subject but also and independently in the other successful subjects taken collectively. The latter finding makes it difficult to sustain any explanation of the effect in terms of a psychological peculiarity of an individual subject and therefore reduces the plausibility of the main counterhypothesis suggested by Pratt and Woodruff.**

**Two further counterhypotheses advanced by Pratt and Woodruff are shown not to account for Hansel's observation.**

**The present analysis therefore tends to support Hansel's interpretation of the experiment.**

## INTRODUCTION

The Pratt-Woodruff experiment (Pratt & Woodruff, 1939) consisted of a lengthy series of card-guessing tests under "clairvoyance" conditions, carried out in 1938-39. Sixty-six subjects took part and 3,868 runs, each of 25 trials, were completed. The whole series divides chronologically into two sections, A and B. Series A may be regarded as of an introductory nature, no very rigorous controls being imposed. Series B, however, is distinguished by the introduction of stringent control conditions. This part of the experiment comprised 2,400 runs, each of 25 trials, carried out by 32 subjects.

As planned, the experiment was intended as an investigation of the effect of symbol size and shape on the scoring rate. As it turned out, no variation of scoring rate with the size of symbol was found. The overall score, however, was highly significant, and the work

[1] The authors wish to thank the Institute for Parapsychology for supplying the thermofax copies of the original score sheets on which the analyses reported in this paper are based.

*Ed's Note*—Dr. Scott is a statistician working in London for UNESCO and the International Statistical Institute. Dr. Medhurst, at the time of his death in 1971, was a mathematician with the General Electric Company in London.

came to be cited as one of the key experiments rigorously demonstrating the existence of ESP. Thus, Rhine and Pratt (1957) remark:

> Those who wish to acquire a reading acquaintance with the highest standards of controlled psi testing may, for example, consult the Pratt and Woodruff report. [p. 39]

In the same reference they add:

> Through the succeeding years [i.e., following the Pearce-Pratt Series, 1933-34] a number of other experiments followed in which the standards of control required for verification were maintained. Perhaps the most elaborately controlled of these was that published by Pratt and Woodruff in 1939 . . . . the scoring rate was highly significant and chance as well as all other conceivable hypotheses were ruled out, leaving only the hypothesis of ESP. [p. 47]

One of the "other conceivable hypotheses" discussed (Pratt & Woodruff, 1939) was that of fraud by the experimenters. Pratt and Woodruff remark:

> Experimental conditions which would make it impossible for one investigator [i.e., working on his own] willfully to deceive his colleagues might not be attainable. However, it is worth pointing out that the conditions of Series B accomplished something in this direction, inasmuch as they made it difficult, if not out of the question, for one experimenter to practice deception upon the other even if he had wished to do so. [p. 140]

Rhine et al. (1940) put the same point rather more strongly:

> . . . these series, Pearce-Pratt, Warner, Pratt and Price, Pratt and Woodruff, and others, also offer difficulty for the hypothesis of untrustworthy investigators. They could be explained by this hypothesis only by supposing *collusion between the two experimenters.* [p. 148] [Italics are those of the original.]

In 1961, Prof. C. E. M. Hansel suggested a possible method by which the successful score in the important Series B test could have been produced improperly by one of the experimenters without the knowledge of the other (Hansel, 1961). He stated that in an experimental reproduction of the Pratt-Woodruff procedure he had found the method to be workable. This in itself might be thought to lessen the claim of the Pratt-Woodruff experiment to be regarded as one of the key results establishing the reality of ESP. But Hansel went

much further. He claimed to have established that the distribution of hits showed a very peculiar feature readily interpreted according to his hypothesis but "difficult to explain in terms of any other hypothesis including that of ESP." He admitted that "the analyses made in this paper are by no means exhaustive, nor as complete as is desirable. They were limited by the time available with the records."

In a reply published with Hansel's paper, Dr. J. G. Pratt and Dr. J. L. Woodruff (1961) asserted that Hansel had used fallacious methods of analysis and had also shown a measure of bad faith. ("His actions appear to represent a deliberate attempt to discredit parapsychology by any means [p. 115].") They suggested, as an argument against the fraud hypothesis, that Woodruff had no more reason for being a "conscious cheat" in this than in any other "psi research reports" with which he was concerned.[2] Further, while admitting that the effect claimed by Hansel could be demonstrated in the scoring pattern of the subject P.M. (the most successful of the 32 subjects who took part in Series B), Pratt and Woodruff maintained that, contrary to Hansel's assertion, this effect could not be shown for the other high-scoring subjects. Their comment on this is as follows:

> Searching through the work of the highest scoring subject, he [Hansel] came upon something which he could interpret as evidence of fraud by Woodruff. Because of this groping, "after-the-fact" approach, this initial finding could not be conclusive even if the method of analysis had been adequate.[3] Confirmation of the effect in the data of other high-scoring subjects was therefore of paramount importance. Hansel's efforts to achieve this objective show that he recognized this

[2] Hansel (1966) has countered this as follows: "In the case of each of the major experimental investigations to which a chapter has been given in this book, there is a possible monetary or prestige motive for trickery. . . . The Pratt-Woodruff experiment was a continuation of work started by Woodruff constituting part of the requirement for a higher degree [p. 235]." In fact, the subject of Dr. Woodruff's M.A. thesis is the research reported in the Pratt-Woodruff article of 1939.

[3] It is not clear on what evidence the authors base this description of how Hansel made his discovery. Actually, the Hansel effect is so *recherché* that it seems unlikely to have emerged from an undirected search. Moreover, it follows fairly directly from Hansel's fraud hypothesis which is, in turn, presumably one of a very small number of non-ESP hypotheses consistent with the reported experimental conditions. Thus Hansel's discovery seems nearer to a *confirmation of a prediction* than to an effect revealed by a groping search.

need. These efforts failed—as we have shown—in spite of his claims to the contrary. [p. 124]

Having apparently established that the "Hansel effect" only exists in one subject, it became open to Pratt and Woodruff to offer an explanation involving individual psychological peculiarities of this subject. Alternatively, they suggested that "this may be a selected, meaningless, statistical effect, for statistical oddities are a dime a dozen." However, in view of the significance level involved ($2 \times 10^{-6}$ for the subject P.M.) it is hard to take the latter proposition seriously.

As will become clear when we have described in detail the Pratt-Woodruff experiment and Hansel's hypothesis, one is compelled to agree with Dr. Pratt and Dr. Woodruff that "confirmation of the effect in the data of other high-scoring subjects [is] . . . of paramount importance." The essentially new material in the present paper is mainly concerned with these subjects.

Before concluding this introduction something should perhaps be said about the propriety of such probing into history. The usefulness of detailed examination of past experiments has sometimes been questioned, partly perhaps because preoccupation with the past is not usually met with in the physical sciences. But parapsychology is distinguished from other fields of research by the peculiarity that no new experiment, however negative its result, is accepted as evidence *against* the phenomenon. Such a situation provides the ideal ground in which erroneous beliefs can take root, and periodic reexamination of the old research takes on a special importance as the only remaining safeguard against this eventuality. An allied viewpoint has been expressed by Dr. Scriven (1961; see also Woodruff, 1961) when he draws attention to the "vulnerability of the key work." Enlarging on this he says:

> By "key work" I mean the few experiments with overwhelmingly positive results. There are good reasons for assessing the evidential value of these as higher than that of many less striking experiments even when the combined mathematical probability from the latter is the same as from the key work. . . . Now there has simply not been enough of these key experiments *in recent years* to stifle a feeling of uneasiness in many of us. . . . [Hansel] has exposed an Achilles heel in the data that we had not previously fully recognized. It is too highly

dependent on too small a family of key successes. The effect of this is to make it too susceptible to being explained away by a single counter-hypothesis, whether it involves fraud or not. [p. 309]

## The Pratt-Woodruff Procedure

We may confine attention to the Series B tests, since these form that part of the experiment in which a serious attempt was made to set up fraud-proof conditions. Photographs and drawings in the original Pratt-Woodruff article (1939) and the book by Rhine et al. (1940) give quite a clear picture of the experimental arrangements. Only an outline need be given here.

During each run, the subject and Experimenter 1 (Woodruff) sat on opposite sides of a table, separated by a screen which extended about up to eye level. A horizontal slot was cut in the bottom of the screen so that an area of table on the subject's side was visible to Experimenter 1. A sloping board was so disposed as to shield from the subject an area of table on the side of the experimenter. (See Figures 1 and 2. These diagrams are not to scale.) On the subject's side of the screen were five pegs in a row over the slot, on which could be hung five cards (described as "key cards") each bearing one of the five symbols: cross, star, circle, square, wavy lines. Near to the slot, on the subject's side of the table, a row of five blank cards were placed flat on the table, one under each peg. These blank cards were visible to Experimenter 1 over the top of the sloping board. The experimenter was provided with the usual pack of 25 ESP cards consisting of five of each of the five symbols. This pack was carefully shuffled before each run.

At the beginning of a run, Experimenter 2 (Pratt) handed the five key cards to the subject, and the latter placed them, in some order chosen by himself, on the pegs. According to the account in Pratt and Woodruff (1939), the first experimenter, holding his pack face downwards, took cards off the pack one by one, while the subject simultaneously touched with a pointer the blank card under the key card whose symbol, so the subject believed, corresponded to that on the card being handled by the experimenter. Then, without looking at the face of the card, Experimenter 1 is stated to have placed this card in one of five piles corresponding to the position of the blank card indicated by the subject. Throughout the run, all the

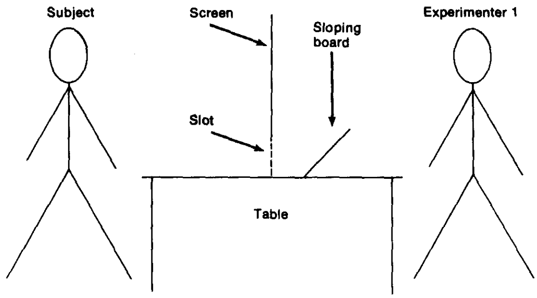


FIG. 1. Positions of subject and Experimenter 1.

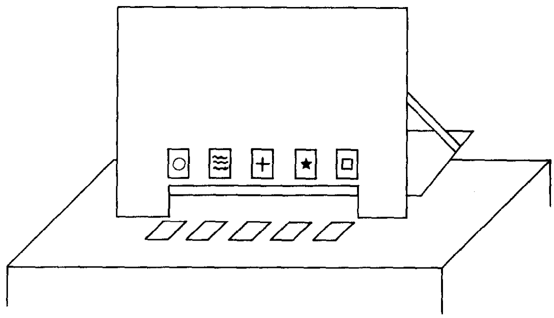

FIG. 2. Screen as seen by subject.

cards handled by Experimenter 1 were hidden from the subject by the sloping board. When the run was completed, the first experimenter recorded the card symbols in each pile on a numbered record sheet while Experimenter 2 recorded the positions of the key cards on a record sheet bearing the same number. These sheets were later compared by a third person when checking for hits. The number of hits was also checked immediately after the run by laying the screen on its side so that the key cards and the 25 target cards were visible together, Experimenter 2 then sorting out and counting the hits from each pile. The screen was now replaced, and Experimenter 2 removed the key cards and handed them to the subject, who replaced them on the pegs, usually in a different order. Experimenter 1 shuffled his pack, Experimenter 2 returned to his seat, and the next run commenced.

This procedure was used for 2,000 of the 2,400 runs. These were described as "STM" (screened touch matching) trials. A variation

was introduced into the remaining 400 runs. During these runs, described as "BSTM" (blind screened touch matching) trials, the key cards were hung up by Experimenter 2 with their backs towards the subject. The intention with this latter procedure was that, supposing there were no identifying marks on the card backs, the subject would be unaware of the order of the cards. Successful scoring could thus be taken as indicating a sort of double clairvoyance. The average scoring rate over all subjects who took part in the BSTM trials was closely similar to that during the main run of STM trials.

The time for each run, including the checking, averaged about two minutes. This, as can be readily verified, requires quite rapid and efficient execution of the actions described. Any hypothesis of fraud must take account of the very short time available for the necessary additional activity.

Throughout the runs, Experimenter 2 sat so that the subject's actions were visible to him, but the first experimenter's hands, and his pack, were completely shielded. One member (Sells, 1939) of the review committee which refereed the Pratt-Woodruff paper did, in fact, ask, "Why didn't the second E [experimenter] watch the first E instead of the S [subject] alone?" This important question received, at the time, no answer.

The overall score was found to be highly significant. A positive deviation of 489 hits was secured, corresponding to a critical ratio of 4.99 (probability $3 \times 10^{-7}$). Most of the significance depends on the trials carried out with the subject P.M. During 162 runs, she produced a positive deviation of 136 hits, corresponding to a critical ratio of 5.34 (probability $4.6 \times 10^{-8}$). The results obtained by the five subjects whose scores reached the highest significance level are shown in Table 1. This table includes both STM and BSTM runs.

It is of interest to see what is the probability level for the Series B data after removal, in turn, of the trials of the successive high-scoring subjects in order of their significance level. Table 2 has been derived from the table in Pratt and Woodruff (1939). Any small discrepancies, such as noted above in connection with subject C.C., will not alter the general pattern. It is evident from this table that, after the removal from the total of the data relating to the five highest-scoring subjects, there is little residual "significance" to account for.

Table 1
TOTAL SCORES OF EACH OF THE FIVE SUBJECTS SHOWING THE HIGHEST SIGNIFICANCE LEVELS

| Subject | No. of Runs Each of 25 Trials | Deviation of Hits from Chance Expectation | Average No. Hits per 25 Trials | *CR* | *P* (One-tailed) |
|---|---|---|---|---|---|
| P.M. | 162 | +136 | 5.84 | 5.34 | $4.6 \times 10^{-8}$ |
| D.A. | 50 | + 43 | 5.86 | 3.04 | $1.2 \times 10^{-3}$ |
| H.G. | 187 | + 76 | 5.41 | 2.78 | $2.7 \times 10^{-3}$ |
| C.C. | 195 | + 72 | 5.37 | 2.58 | $4.9 \times 10^{-3}$ |
| D.L. | 119 | + 46 | 5.39 | 2.11 | $1.7 \times 10^{-2}$ |

*Notes.* 1. The figures given for subject CC are slightly different from those shown in Tables II and IIB of Pratt and Woodruff (1939). A check of copies of the score sheets reveals that for this subject Pratt and Woodruff omit one run with a score of 2.

2. Probabilities quoted are one-tailed as in Table IIB of Pratt and Woodruff (1939); that is, they are the probabilities of getting the observed or a larger positive deviation.

Table 2
RESIDUAL SIGNIFICANCE AFTER REMOVAL OF DATA FOR THE HIGHEST-SCORING SUBJECTS FROM TOTAL SERIES B DATA

| Subjects Subtracted from Total | No. of Runs Each of 25 Trials | Deviation of Hits from Chance Expectation | Average No. Hits per 25 Trials | *CR* | *P* (One-tailed) |
|---|---|---|---|---|---|
| None (pooled trials of all subjects) | 2,400 | +489 | 5.204 | 4.991 | $3.0 \times 10^{-7}$ |
| P.M. | 2,238 | +353 | 5.158 | 3.731 | $9.5 \times 10^{-5}$ |
| P.M.+D.A. | 2,188 | +310 | 5.142 | 3.314 | $4.6 \times 10^{-4}$ |
| P.M.+D.A.+H.G. | 2,001 | +234 | 5.117 | 2.616 | $4.4 \times 10^{-3}$ |
| P.M.+D.A.+H.G.+ C.C. | 1,807 | +162 | 5.090 | 1.905 | $2.8 \times 10^{-2}$ |
| P.M.+D.A.+H.G.+ C.C.+D.L. | 1,688 | +116 | 5.069 | 1.412 | $7.9 \times 10^{-2}$ |

## HANSEL'S HYPOTHESIS

Hansel points out that at the end of a run, when the board was laid on its side during the checking, Experimenter 1 became aware of the order of the key cards in the run just concluded. Before the

next run the board was put back into its normal upright position, the key cards being taken off their pegs by Experimenter 2, presumably in order, and handed to the subject, who replaced them, usually, in a different order. If Experimenter 2 had shuffled the key cards while they were in his hands, the first experimenter's knowledge of their order during the previous run would have been irrelevant. However, so far as the published evidence goes, shuffling did not take place, and there appears to be no reason why, at the time, it would have been thought to be necessary. This point is discussed further in the following section.

It thus appears that we may fairly take as a hypothesis that the order of the key cards was known or knowable to Experimenter 1, at least on many occasions, when they were handed to the subject. It appears from the photographs in the Pratt-Woodruff article (1939) and the Rhine-Pratt book (1957) that the experimenter could see the subject's eyes over the screen, and perhaps in addition (though this is not directly verifiable) the shadow of her arm in the region of the table visible through the slot in the screen. Hansel suggests that sufficient visual clues of this sort were available for Experimenter 1 to infer on which pegs successive cards were hung. The range of eye movements, etc., as the cards were put into position, may be expected to have been by no means negligible, since the distance between extreme pegs appears to have been fifteen inches or more.[4]

This knowledge would have enabled a significant score to be produced artificially if one or more target cards, during a run, were placed by the experimenter, not in the position indicated by the subject but in the position producing a correspondence between target and key card. It should be noticed that even for the highest-scoring subject, the hypothetical misplacement would be needed, on average, only once per run to account for the observed result.

It might be thought that the procedure just described would have

[4] Pratt and Woodruff (1961) claim that the same information would not have been available to Experimenter 1 during the BSTM tests, but the reason for this assertion is not obvious. It seems necessary only to substitute "Experimenter 2" for "the subject" in the foregoing. Incidentally, BSTM runs are not distinguished on the score sheets and are not identified in the report. We have therefore included them without distinction in our analyses.

called for an unlikely degree of skill on the part of Experimenter 1. However, Hansel (1961) claims to have verified by experiment that it can be carried out with no great difficulty. He remarked:

> I found it far simpler to obtain high scores when acting as experimenter by noting the symbol occupying one of the end positions and detecting the position at which it was replaced on the pegs. I instructed an assistant to take the key cards from the pegs in order from left to right and then to replace them in different positions on the pegs. It was then quite easy to note the position at which the cards were replaced. This was done by listening to the sound of the cards being replaced and by observing the shadow of his arm on the table under the slit at the bottom of the screen. The positions of the first and last cards replaced were easiest to identify in this way. [p. 104]

The statistical analysis presented in Hansel's paper is based on this latter observation.

We have repeated this test ourselves and can confirm that the manipulation is easily performed. We did not find the position of the cards placed first or last any easier to identify than the remainder. However, a factor which certainly favors the first and last cards is the difficulty of *memorizing* the card order of the previous run. Only one or two cards can be easily recalled and it seems natural to concentrate on remembering one or both of the end cards.

We wish to make it clear that the hypothesis of card misplacement by the experimenter does not necessarily imply conscious deception. The target cards appear to have been used many times over and might have become identifiable to the experimenter by their backs. He might then occasionally misplace one unconsciously in the pile where he knew it "ought" to go. In the rest of this paper we refer to the "card misplacement hypothesis" originated by Hansel without wishing to imply that the misplacement was necessarily intentional.

## Evidence of the Shuffling of the Key Cards

As already remarked, if Experimenter 2 had shuffled the key cards, however cursorily, after the checking and before handing them to the subject, the information required to operate Hansel's misplacement procedure would have been destroyed. No mention of shuffling the key cards appears in Pratt and Woodruff (1939) and

Rhine et al. (1940), and it seems clear that at the time of the experiment shuffling would not have been felt to be necessary.

Since this is clearly a vital point, Hansel tried to resolve the question by showing that for subjects P.M. and D.A. there is a very significant correlation between the orders of the key cards in successive runs. This, he argues, shows that shuffling could not have occurred. However, in the case of P.M. the correlation might well be due to inefficient shuffling.

Dr. Pratt, in conversation with one of us (R.G.M.), has stated that, according to his recollection, the key cards *were* shuffled before each run, and he told Hansel the same thing (Hansel, 1961). If this were so, it would appear that more than one subject had a habitual tendency to carry over memory of the key card arrangement from run to run, and to replace the cards in a similar order irrespective of the order in which they were presented. This is clear from the score sheets of subjects D.L. and D.A. The same key-card sequence appears in successive runs 6 times out of 111 in the former and 11 times out of 47 in the latter, both results being highly significant on a binomial test (i.e., on the assumption of complete randomization between runs). On balance, in view of the irrelevance of shuffling to the declared purpose of the experiment, the absence of a definite statement in the original report, and the correlation between successive key-card arrangements, it seems likely to us that Dr. Pratt's recollection is in this respect faulty (no unlikely thing in regard to a minor detail of a twenty-year-old sequence of events).

## Hansel's Analysis

If card misplacement took place on the lines postulated by Hansel, nothing statistically unusual would appear in the scoring pattern if Experimenter 1 favored all key-card positions impartially. However, Hansel's experimental observation that the key cards on the extreme left or right were easiest to keep track of between runs suggests that it would be worth while to see whether there is a significant excess of hits in the piles of target cards whose associated key cards occupied end positions in the *previous* run. We shall call these the E-piles, and the remaining piles the M-piles. Hansel was successful in demonstrating this effect in the scoring pattern of the highest-scoring subject,

P.M. However, his claim to have done likewise for other high-scoring subjects was easily refuted by Pratt and Woodruff, since he had, as they remark, "loaded the results in favor of his prediction by re-using the P.M. data."[5]

Another weakness of Hansel's analysis is that it was confined to those runs in which the score was above 5 and in which no two or more piles showed the same highest score. This is an unnecessary restriction. The supposed misplacement could have occurred in runs in which the total score happened to be 5 or less because of a low number of chance hits among the trials not tampered with.

## Distribution of Hits between the E- and M-Piles for the Five Highest-Scoring Subjects

In Table 3, all the data (including both STM and BSTM runs) have been taken for each subject except for the omission of the first run of each session, in which the postulated misplacement could not have occurred. The probability levels (one-tailed) shown for each set of data are based on a chance expectation of 5 hits per 25 trials. The totalled hits and trials shown in this table differ, of course, from those in Table 1 since the latter includes all runs for each subject.

The consistently better performance on the E-piles is striking, and particularly so for P.M., where the whole of a very high "significance" is concentrated in the E-piles. The scoring pattern appears consistent with the misplacement postulated by Hansel. However, the question that has to be asked, and which is the primary subject of this paper, is whether, on the assumption of extrasensory perception, the observed higher proportions of hits in the E-piles could reasonably be said to have occurred by chance, which for the subjects other than P.M. is what Pratt and Woodruff claim.[6]

Before considering this, one further observation may be of interest. If Hansel's postulated misplacement were to occur, it might well be expected that the substitution would be made on the first

[5] Subsequently, in his book, Hansel (1966) offered a separate analysis for the other high-scoring subjects. However, there are errors in the data he presented (in Table 5), and we have not been able to confirm the significance level that he reported.

[6] Detailed results by subject and pile position for the five subjects are shown in the appendix to this paper.

occasion that the desired target card made its appearance. If this were done, a high concentration of hits would be expected at the *bottoms* of the E-piles. On the score sheets, each pile of target cards, in each run, is recorded in a vertical column. The account of the checking procedure is not sufficiently clear to make it certain whether the bottom card of a pile corresponds to the top or to the bottom entry of a column on the score sheet. However, it presumably corresponds to one or the other, so that it becomes of interest to further subdivide the data by separating off the top and bottom target cards in the E-piles and comparing the proportion of hits on these with the remainder of the data. Table 4 shows the result of this operation.[7]

Table 3

COMPARISON OF PERFORMANCE ON THE E- AND M-PILES OF THE FIVE HIGHEST SCORING SUBJECTS

| Subject | E-Piles | | | | | M-Piles | | | | | *P* of Total Score (from Table 1) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Trials | Hits | Hits per 25 Trials | *CR* | *P* (One-tailed) | Trials | Hits | Hits per 25 Trials | *CR* | *P* (One-tailed) | |
| P.M. | 1,600 | 434 | 6.78 | 7.13 | $5.2 \times 10^{-13}$ | 2,249 | 465 | 5.17 | 0.79 | $2.1 \times 10^{-1}$ | $4.6 \times 10^{-8}$ |
| D.A. | 473 | 124 | 6.55 | 3.38 | $3.6 \times 10^{-4}$ | 701 | 156 | 5.56 | 1.49 | $6.8 \times 10^{-2}$ | $1.2 \times 10^{-3}$ |
| H.G. | 1,764 | 396 | 5.61 | 2.57 | $5.1 \times 10^{-3}$ | 2,687 | 562 | 5.23 | 1.20 | $1.2 \times 10^{-1}$ | $2.7 \times 10^{-3}$ |
| C.C. | 1,880 | 410 | 5.46 | 1.97 | $2.4 \times 10^{-2}$ | 2,794 | 595 | 5.32 | 1.69 | $4.5 \times 10^{-2}$ | $4.9 \times 10^{-3}$ |
| D.L. | 1,135 | 265 | 5.84 | 2.82 | $2.4 \times 10^{-3}$ | 1,639 | 335 | 5.11 | 0.44 | $3.3 \times 10^{-1}$ | $1.7 \times 10^{-2}$ |

*Note.* It will be seen that the total trials shown for each subject are not exact multiples of 25. For P.M., D.A., C.C. and D.L., only 24 trials are entered on the score sheet for one of the runs. For H.G., 26 trials are entered for one of the runs.

Comparing Table 4 with Table 3, it is seen that the further narrowing down of the period during each run when the suggested misplacement seems most likely to have occurred has, in the case of each subject, caused an increase in the average scoring rate for the "suspect" trials. This observation does not argue differentially in favor of Hansel's hypothesis and against the ESP hypothesis, since "salience" (concentration of scoring at the ends of runs) has been widely

[7] Pratt (1961) has drawn attention to this variation in the scoring rate in relation to position in the run (without, of course, distinguishing E- and M-piles), which he regards as a salience effect.

reported as an ESP effect. However, it is of some interest that this phenomenon, which might be expected on Hansel's hypothesis, is in fact observed in a marked form.

Table 4
COMPARISON OF PERFORMANCE ON TRIALS AT TOPS AND BOTTOMS OF E-PILES WITH REMAINING TRIALS

| Subject | Trials at Tops and Bottoms of E-Piles | | | All Other Trials | | |
|---|---|---|---|---|---|---|
| | Trials | Hits | Average Hits per 25 Trials | Trials | Hits | Average Hits per 25 Trials |
| P.M. | 616 | 197 | 8.00 | 3,233 | 702 | 5.43 |
| D.A. | 188 | 52 | 6.91 | 986 | 228 | 5.78 |
| H.G. | 712 | 178 | 6.25 | 3,739 | 780 | 5.22 |
| C.C. | 748 | 171 | 5.72 | 3,926 | 834 | 5.31 |
| D.L. | 444 | 112 | 6.31 | 2,330 | 488 | 5.24 |

## STATISTICAL SIGNIFICANCE OF THE EXCESS OF HITS IN THE E-PILES

It has been shown in the preceding section that in the scores of each of the five highest-scoring subjects there is an excess of hits in the E-piles as compared with the M-piles. The question now to be considered is whether these excesses are statistically significant; that is, we have to ask with what probability would the observed excesses have occurred by chance, supposing the ESP were operating to produce the observed scoring level for each subject.[8]

Pratt and Woodruff do not deny the existence of a highly significant "Hansel effect" for subject P.M., though their interpretation is not that of Hansel. Since neither Pratt and Woodruff nor Hansel used the whole of the relevant data in their analyses, even for the trials with P.M., it will be of interest, before considering the other subjects, briefly to review the P.M. trials.

[8] Note that we are testing for an effect compatible with Hansel's postulated "misplacement" when operated in the way which he himself found easiest. Failure to find a statistically significant difference between E- and M-pile scores would not refute Hansel's general hypothesis, though success in such a search would lend support to it.

*Trials with Subject P.M.*

Table 5 shows the relevant data for P.M.

The value of $\chi^2$, corrected for continuity, is 21.36, and the corresponding probability (one-tailed) is $1.9 \times 10^{-6}$. There can thus be no reasonable grounds for treating the division between the two groups of trials as a chance event.

Table 5
CONTINGENCY TABLE FOR SUBJECT P.M.

| | Hits | Misses | Total |
|---|---|---|---|
| E-Piles | 434 | 1,166 | 1,600 |
| M-Piles | 465 | 1,784 | 2,249 |
| Total | 899 | 2,950 | 3,849 |

*Trials with Subjects D.A., H.G., C.C., D.L.*

The contingency table for the pooled data for these subjects is shown in Table 6.

Table 6
CONTINGENCY TABLE FOR SUBJECTS D.A., H.G., C.C., D.L. POOLED

| | Hits | Misses | Total |
|---|---|---|---|
| E-Piles | 1,195 | 4,057 | 5,252 |
| M-Piles | 1,648 | 6,173 | 7,821 |
| Total | 2,843 | 10,230 | 13,073 |

The corrected $\chi^2$ is 5.12 and the corresponding probability (one-tailed) is 0.012.[9]

The single-tailed probability is appropriate for these tests because

[9] The Fisher exact test (Robertson, 1960) gives a probability of 0.0119. It could be argued that instead of pooling subjects who have different scoring rates, we should compute the expectations and variance for each of the four subjects separately and sum them before computing $\chi^2$. For justification of this procedure, see Cochran (1954), Mantel and Haenszel (1959), Mantel (1963), and Birch (1964). If this is done we obtain a probability of 0.0119 again.

we are testing not for *any* type of uneven distribution between the groups of trials but specifically for an *excess* in the E-piles, such a tendency to excess being both predicted by Hansel's experimental findings and established for the highest-scoring subject, P.M. Thus there is significant evidence of a "Hansel effect" in subjects other than P.M.

The reader may wonder how it is that Pratt and Woodruff (1961), testing broadly the same hypothesis as we, arrive at a non-significant result. There are three main reasons for this:

1. Pratt and Woodruff use a two-tailed test. We have already explained why this is inappropriate.
2. Pratt and Woodruff test a $2 \times 5$ instead of a $2 \times 2$ contingency table. This test is sensitive to types of departure from the null hypothesis not predicted by Hansel's hypothesis, which makes it relatively insensitive to departures of the type predicted by Hansel and which we wish to test. (Adoption of a test that is too broad, i.e., that is sensitive to phenomena other than those of interest, weakens the power of the test and may lead to failure to obtain a significant result for the phenomena of interest.)
3. Pratt and Woodruff test only subjects H.G. and C.C. while we have included all subjects who individually achieved a probability below 0.05 in the main "ESP effect." Some selection of subjects is necessary since clearly the Hansel hypothesis makes no prediction about subjects whose scores are in accordance with chance. The 0.05 criterion is hallowed by tradition and we believe that by following it we have gone as far as humanly possible to protect ourselves against any suspicion of selection of data after the fact.

Pratt and Woodruff also (with Hansel) limit their analyses to runs scoring 6 or more. This raises a more difficult selection problem. While it is true that the factor making for successful card-calling should be more highly operative in the high-scoring runs, nevertheless it should be present to some extent even in the below-expectation runs. It is not at all clear where to draw the line and again there is danger of selection after the fact. We decided to include all runs. This decision entails some risk of diluting the effect, but in fact we turn out to have enough significance left, and at least

this seems to offer complete protection from the critic suspicious of motivated selection.

## Alternative Hypotheses

Is there any other "innocent" hypothesis which might explain the uneven distribution of hits between E- and M-piles without recourse to Hansel's theory of fraud? Two suggestions advanced by Pratt and Woodruff (1961) seem worth following up. The first is described as follows:

> Usually, subjects show a tendency to respond more often[10] to the symbols occupying the three inner key-card positions. If there exists at the same time a tendency for subjects to place the key cards that were in the end positions back on the three inner pegs, these two concomitant habits would account for finding *more hits* on the symbols that had been on the ends. Thus this hypothesis might explain Hansel's results without the necessity of bringing in either his trickery interpretation or any variation of the ESP hypothesis. [pp. 119-20]

We shall refer to this as the "position preference hypothesis."

A second hypothesis is advanced in the same reference:

> There may be an alternative ESP interpretation, such as a differential rate of scoring on the five symbols coupled with some habitual tendency in the placement of the symbols on the pegs. [p. 126]

This is similar to the position preference hypothesis except that the *identity of the symbol* now plays the role of the *position in the current run*. We shall refer to this as the "symbol preference hypothesis."

Both hypotheses are advanced purely speculatively by Pratt and Woodruff, though it is easy to check them (easy, that is, in principle but laborious in practice). For example, to test the symbol preference hypothesis we determine, for each subject separately, the scoring rate (i.e., proportion of hits) for each symbol called. The calls on each symbol can be classed as E or M depending on whether that symbol was at the *end* of the row in the previous run or in one of the

[10] "Respond more often" here might mean either to "call more often" or to "call successfully more often"; i.e., to score at a higher rate. If the former, the hypothesis would not explain the significant result in the chi-square test applied above to Tables 5 and 6, which compares the distribution of hits with that of misses. We therefore assume that the hypothesis refers to a possible tendency to score a *higher proportion of hits* on the three inner card positions.

*middle* positions. For each symbol, we count the number of E's and M's among the calls and multiply these by the scoring rate for that symbol. This gives the *expected* score on E- and M- calls allowing for symbol preference. This is done separately for each subject. Summation then gives the total expected E- and M-hits on the symbol preference hypothesis. The observed number of E- and M-hits is then compared with the expected by a contingency test. An exactly analogous procedure deals with the position preference hypothesis.

Results are shown in Table 7. It is immediately clear that neither hypothesis accounts for the observed results.

Table 7
SIGNIFICANCE OF EXCESS HITS ON E-SYMBOLS ON THREE HYPOTHESES

| | Subject P.M. | | | Subjects D.A., H.G., C.C., D.L. | | |
|---|---|---|---|---|---|---|
| | No. Hits on E-Symbols | $\chi^2$ (1 *df*) | *P* | No. Hits on E-Symbols | $\chi^2$ (1 *df*) | *P* |
| Observed | 434 | ........ | ................ | 1,195 | ........ | ................ |
| Expected: | | | | | | |
| Unadjusted | 373.71 | 21.36 | $1.9 \times 10^{-6}$ | 1,142.16 | 5.12 | $1.2 \times 10^{-2}$ |
| Allowing for position preference | 373.26 | 21.69 | $1.6 \times 10^{-6}$ | 1,142.49 | 5.06 | $1.2 \times 10^{-2}$ |
| Allowing for symbol preference | 377.73 | 18.55 | $8.4 \times 10^{-6}$ | 1,147.87 | 4.06 | $2.2 \times 10^{-2}$ |

*Note.* Each of the 6 chi-squares is based on a $2 \times 2$ table (hits/misses : E/M). Only one of the 4 cells is shown in each case, the other 3 being deducible from the marginal totals, which are the same as in Tables 5 and 6. The "unadjusted" line gives Tables 5 and 6, in which the expectations are computed from the marginal totals on the usual proportional basis. For the last two lines, the expectations are computed as described above. Yates's correction is used throughout and probabilities are one-tailed.

## CONCLUSION AND COMMENTS

The probability level (0.012) found for the excess of E-pile hits in the trials of the four highest-scoring subjects other than P.M.

would be taken, in general scientific practice, as evidence of a real, non-fortuitous effect. In coming to such a decision it would be borne in mind (a) that we are looking for confirmation of an effect already established for subject P.M., and (b) that each subject separately shows the predicted excess.

It is still open, for those who find it easier to accept the reality of clairvoyance than Hansel's hypothesis of card misplacement by an experimenter, to say, as do Pratt and Woodruff (1961) that "statistical oddities are a dime a dozen." However, much of the literature of parapsychology has been taken up with refutations of just this argument when it is advanced by those skeptical of the evidence for paranormal effects. Dr. Pratt (1964) has remarked:

> When the chance odds at which we arrive by our statistics are as unlikely as 1 in 100 [i.e., probability 0.01], scientists generally agree that it is not reasonable to say that chance alone was involved. Therefore we reject chance and look for some lawful principle at work in the experiment. [p. 51]

Pratt and Woodruff (1961) offered a "consistent and reasonable ESP hypothesis" for the subject P.M. in terms of her individual psychology, as follows:

> For the subject P.M., the run began, in the psychological sense, when she re-arranged and placed the target cards. The ESP task being a difficult one, she dealt with it by a "narrowing of attention" procedure. For her the task became one of attempting to identify only *some* of the cards in the deck: those with the particular symbols which had become salient because of their prominent, end positions in the preceding run. [p. 126]

It would clearly be far-fetched to offer the same explanation for the effect found in the scoring pattern of the other subjects. If it is accepted that the observed excess of E-pile hits for subjects D.A., H.G., C.C., and D.L. has been shown not to be fortuitous, Hansel's hypothesis seems to be definitely reinforced.

## Summary

The present position regarding the Pratt-Woodruff experiment and Hansel's criticism thereof may be summarized as follows:

1. Hansel suggested a hypothesis of card misplacement by one ex-

perimenter which appears to be consistent with the experimental conditions described in the original report.

2. This misplacement might be carried out in various ways but the apparently easiest way would have certain consequences detectable in the data.
3. Hansel showed these consequences to be present in the data for the highest-scoring subject, and attempted to show the same for other successful subjects.
4. Pratt and Woodruff showed that in the latter attempt Hansel had failed.
5. They also offered an after-the-fact speculation to explain the phenomenon in terms of the particular subject's psychological approach to the task.
6. We have shown that the effect is indeed present in the other successful subjects taken collectively.
7. As the effect is of a most obscure nature we submit that it would be implausible to postulate the same psychological peculiarity among these subjects. Further, the finding of this effect in subjects other than the first practically eliminates any suspicion that it may be a statistical artifact discovered by an after-the-fact "groping" search.
8. Two alternative hypotheses suggested by Pratt and Woodruff to account for Hansel's observation are found not to be consistent with the observational results.
9. We conclude that the evidence tends to support Hansel's hypothesis of card misplacement.
10. The evidence is not, of course, compelling. It is open to anyone to prefer the hypothesis that an unlikely coincidence has occurred or that the psychological peculiarity attributed by Pratt and Woodruff to the subject P.M. applied to more than one subject, or to produce yet another hypothesis in terms of an ESP effect. Exactly where the balance of probability lies, in the light of all the evidence, must be, as always, to some extent a matter of opinion, depending among other things on the degree of probability one attaches to the occurrence of various types of experimental error. However, it seems clear that the new evidence in this paper moves the balance at least some distance toward Hansel's hypothesis.

## APPENDIX

### Details of Scoring According to Pile Position for the Five Subjects with Highest Significance Level

| Subject | | Pile Number | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| P.M. | Number of hits | 201 | 147 | 155 | 163 | 233 |
| | Number of trials | 806 | 721 | 753 | 775 | 794 |
| | Average hits per 25 trials | 6.23 | 5.10 | 5.15 | 5.26 | 7.34 |
| D.A. | Number of hits | 59 | 58 | 48 | 50 | 65 |
| | Number of trials | 231 | 227 | 234 | 240 | 242 |
| | Average hits per 25 trials | 6.39 | 6.39 | 5.13 | 5.21 | 6.71 |
| H.G. | Number of hits | 210 | 194 | 191 | 177 | 186 |
| | Number of trials | 888 | 909 | 885 | 893 | 876 |
| | Average hits per 25 trials | 5.91 | 5.34 | 5.40 | 4.96 | 5.31 |
| C.C. | Number of hits | 204 | 194 | 197 | 204 | 206 |
| | Number of trials | 957 | 938 | 923 | 933 | 923 |
| | Average hits per 25 trials | 5.33 | 5.17 | 5.34 | 5.47 | 5.58 |
| D.L. | Number of hits | 130 | 104 | 117 | 114 | 135 |
| | Number of trials | 575 | 540 | 556 | 543 | 560 |
| | Average hits per 25 trials | 5.65 | 4.81 | 5.26 | 5.25 | 6.03 |
| D.A.+H.G.+ C.C.+D.L. | Number of hits | 603 | 550 | 553 | 545 | 592 |
| | Number of trials | 2,651 | 2,614 | 2,598 | 2,609 | 2,601 |
| | Average hits per 25 trials | 5.69 | 5.26 | 5.32 | 5.22 | 5.69 |
| D.A.+H.G.+ C.C.+D.L.+ P.M. | Number of hits | 804 | 697 | 708 | 708 | 825 |
| | Number of trials | 3,457 | 3,335 | 3,351 | 3,384 | 3,395 |
| | Average hits per 25 trials | 5.81 | 5.22 | 5.28 | 5.23 | 6.08 |

## References

Birch, M. W. The detection of partial association: I. The 2 × 2 case. *Journal of the Royal Statistical Society,* 1964, Series B, **26**, 313-24.

Cochran, W. G. Some methods for strengthening the common $\chi^2$ tests. *Biometrics,* 1954, **10**, 417-51.

HANSEL, C. E. M. A critical analysis of the Pratt-Woodruff experiment. *Journal of Parapsychology,* 1961, **25**, 99-113.

HANSEL, C. E. M. *ESP: A Scientific Evaluation.* New York: Charles Scribner's Sons, 1966.

MANTEL, N. Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association,* 1963, **58**, 690-700.

MANTEL, N., & HAENSZEL, W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute,* 1959, **22**, 719-48.

PRATT, J. G. Run salience in the Pratt-Woodruff Series. *Journal of Parapsychology,* 1961, **25**, 130-35.

PRATT, J. G. *Parapsychology: An Insider's View of ESP.* New York: Doubleday, 1964.

PRATT, J. G., & WOODRUFF, J. L. Size of stimulus symbols in extra-sensory perception. *Journal of Parapsychology,* 1939, **3**, 121-58. (Also submitted as M.A. thesis by J. L. Woodruff, 1939.)

PRATT, J. G., & WOODRUFF, J. L. Refutation of Hansel's allegation concerning the Pratt-Woodruff series. *Journal of Parapsychology,* 1961, **25**, 114-29.

RHINE, J. B., & PRATT, J. G. *Parapsychology: Frontier Science of the Mind.* Springfield, Ill.: Charles C Thomas, 1957.

RHINE, J. B.; PRATT, J. G.; SMITH, B. M.; STUART, C. E.; & GREENWOOD, J. A. *Extra-Sensory Perception after Sixty Years.* New York: Henry Holt, 1940.

ROBERTSON, W. H. Programming Fisher's exact method of comparing two percentages. *Technometrics,* 1960, **2**, 103-7.

SCRIVEN, M. New frontiers of the brain. *Journal of Parapsychology,* 1961, **25**, 305-18.

SELLS, S. B. Letter. *Journal of Parapsychology,* 1939, **3**, 246-48.

WOODRUFF, J. L. Letter. *Journal of Parapsychology,* 1961, **25**, 266-68.

*60 Highgate Hill*
*London N. 19, England*